AD-A043 201 MASSACHUSETTS INST OF TECH CAMBRIDGE DEPT OF MATHEMATICS F/G 12/1
ON THE STRICT DETERMINATION OF HYPOTHESIS TESTING GAMES.(U)
JUL 77 R C PERSONS N00014-76-C-0555
UNCLASSIFIED TR-6 NL

AD A043201

12

6

ON THE STRICT DETERMINATION OF HYPOTHESIS TESTING GAMES.

BY

10

RICHARD CLARK PERSONS

9

TECHNICAL REPORT NO. 6

11 22 JULY 22, 1977

12 31p. 14 TR-6

PREPARED UNDER CONTRACT

15 N00014-75-C-0555 (NR-042-331)

FOR THE OFFICE OF NAVAL RESEARCH

DEPARTMENT OF MATHEMATICS
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS

220 021

ON THE STRICT DETERMINATION OF HYPOTHESIS TESTING GAMES

by

Richard Clark Persons

Technical Report No. 6

July 22, 1977

Prepared Under Contract
N00014-75-C-0555 (NR-042-331)
For the Office of Naval Research

DEPARTMENT OF MATHEMATICS
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS

ON THE STRICT DETERMINATION OF HYPOTHESIS TESTING GAMES[1]

by

Richard Clark Persons

Department of Mathematics
Massachusetts Institute of Technology

Section 0. A Rough Statement of the Question

Consider the following game for two players. Player I picks some measure P_{θ_0} from the collection $\{P_\theta: \theta\varepsilon[0,1]\}$. Each P_θ is a probability measure on $[0,1]$ equipped with the usual Borel sigma-field. A point x is then selected at random via P_{θ_0} , the measure chosen by Player I. Player II must then guess whether $\theta_0 \varepsilon [0,\frac{1}{2}]$ or $\theta_0 \varepsilon (\frac{1}{2},1]$. If he is right he wins; if not he loses.

A strategy for Player I is a measure ν on $\Theta = [0,1]$, the set of all possible values of θ . A strategy for Player II is a function $\phi: [0,1] \to \{0,1\}$, where $\phi(x) = 0$ indicates the guess that $\theta_0 \varepsilon [0,\frac{1}{2}]$, and $\phi(x) = 1$ indicates the guess of $\theta_0 \varepsilon (\frac{1}{2},1]$.

The question to be examined in this paper is: if for each Player I strategy, ν , there is a Player II strategy ϕ_ν that will win with probability one, is there a single strategy ϕ that will win with probability one for any strategy ν of Player I?

This question has an application to the issue of Bayesian statistics

---

[1] Prepared with the support of National Science Foundation Grants No. GP-43805 and MPS 75-10376 and ONR Contract N00014-75-C-0555 (NR 042-331).

AMS 1970 subject classifications. Primary 62B20, 62A15.

Key words and phrases. Hypothesis testing, orthogonal classes of measures, strictly determined games, Bayesian.

versus classical statistics. Consider Player I to be nature and Player II to be the statistician. If the omniscient Bayesian statistician (always knowing nature's prior ν), can always make the right decision in a certain situation, can the classical statistician always make the right decision in the same situation?

The above is a special case of a more general question: Does every Hypothesis Testing Game have a value?

Consider the game as above. Player I selects θ, then a point x is selected according to P_θ. Player II sees x and guesses whether $\theta \varepsilon [0,\frac{1}{2}]$ or $\theta \varepsilon (\frac{1}{2},1]$, winning 1 if correct and 0 if not. A strategy of Player I is a probability measure ν on $[0,1]$, the set of all possible values of θ. He selects θ via the measure ν. A strategy for Player II is a function ϕ mapping $[0,1]$, the set of all possible values of x, to $[0,1]$. We let $\phi(x)$ be the probability of guessing that $\theta \varepsilon [0,\frac{1}{2}]$, when x is the observed value.

Define the expected payoff of Player II as

$$M(\nu,\phi) = \iint I_{[0,\frac{1}{2}]}(\theta)\phi(x)P_\theta(dx)\nu(d\theta)$$

$$+ \iint I_{(\frac{1}{2},1]}(\theta)(1-\phi(x))P_\theta(dx)\nu(d\theta) .$$

The upper value of the game is defined as

$$U = \inf_\nu \sup_\phi M(\nu,\phi) .$$

The lower value L is

$$L = \sup_{\phi} \inf_{\nu} M(\nu,\phi) .$$

A game is said to have a value if U equals L .

What is considered in this paper is the case for which U is one. We ask the question whether this implies that L is one also.

Section 1. Introduction and Summary

From this point on, unless stated otherwise, we will assume that $\{P_\theta, \theta\varepsilon[0,1]\}$ is a collection of countably additive probability measures on the unit interval, equipped with the usual Borel sigma-field. Also needed is a measurability condition on $\{P_\theta\}$ as a function of θ . For every Borel subset A of [0,1] , assume that $P_\theta(A)$ is a Borel function of θ with respect to the usual Borel sigma-field on [0,1] . This measurability condition will be called assumption M .

For ease of notation, we make the following definitions:

$$\Theta = [0,1] , \quad \Theta_1 = [0,\tfrac{1}{2}] , \quad \Theta_2 = (\tfrac{1}{2},1]$$

$\mathcal{B}_\Theta$ = the usual Borel sigma-field on Θ ,

$X = [0,1]$, $\mathcal{B}_X$ = the usual Borel sigma-field on X ,

$P_1 = \{P_\theta : \theta \varepsilon \Theta_1\}$, $P_2 = \{P_\theta : \theta \varepsilon \Theta_2\}$.

Of course, X could be any complete separable metric space, and

similarly for Θ. The sets Θ_1 and Θ_2 could be arbitrary disjoint Borel subsets of Θ.

The families P_1 and P_2 can be orthogonal in three ways: pairwise, weak, and strong. Pairwise orthogonality of P_1 and P_2 will mean that every member of P_1 is orthogonal to every member of P_2. Weak orthogonality will mean that for every probability measure ν_1 on Θ_1 and ν_2 on Θ_2 the measures $\int P_\theta \nu_1(d\theta)$ and $\int P_\theta \nu_2(d\theta)$ are orthogonal. Strong orthogonality will mean that there exists a Borel set B which supports every element of P_1 and B^c supports every element of P_2.

The three types of orthogonality will be denoted by "$P_1 \perp (pr) P_2$", "$P_1 \perp (w) P_2$", and "$P_1 \perp (s) P_2$", respectively.

Now that the notation has been introduced, let us mention why the measurability Assumption M is needed. Without this assumption, it is easy to find families P_1 and P_2 that are "weakly" separated but not strongly separated. What follows is such an example.

Let A be some analytic subset of X that is not Borel. (An analytic set is the forward image of a Borel set via a Borel measurable function. See, for example, Kuratowski, Volume I, (1966), or Parthasarathy (1967).) Parameterize $\{P_\theta\}$ in which a way that $P_1 = \{\delta_x : x \in A\}$ and $P_2 = \{\delta_x : x \in A^c\}$. As is shown in Section 2, P_1 is a legitimate family under the measurability assumption, though P_2 is not.

Analytic sets and complementary analytic sets are measurable with respect to the completion of any given Borel measure. (Refer to Kuratowski,

Section 39.) The obvious definition of an average of elements of P_1 is any measure whose completion assigns measure one to the set A . A similar definition would hold for averages of elements of P_2 . Let Q_1 be any average of P_1-measures and let Q_2 be any average of P_2-measures. Since the completion of Q_1 gives A measure one, there is a Borel set $B_1 \subset A$ such that $Q_1(B_1) = 1$, and similarly there is a Borel set $B_2 \subset A^c$ such that $Q_2(B_2) = 1$. It follows that $Q_1 \perp Q_2$ giving $P_1 \perp (w) P_2$. However P_1 and P_2 are not strongly separated by Borel sets since A is not a Borel set.

Assumption M could be weakened in one way suggested by the following example. Let $P_1 = \{r(\theta)\delta_0 + (1-r(\theta))\delta_{1/3} : \theta \varepsilon \Theta_1\}$ and $P_2 = \{r(\theta)\delta_{2/3} + (1-r(\theta))\delta_1 : \theta \varepsilon \Theta_2\}$ where $r : \Theta \to (0,1)$ is any function. If r is a non-measurable function then Assumption M is not satisfied even though the classes P_1 and P_2 are easily separated by a Borel set. If P_1 , P_2 are any two classes of measures generated under Assumption M then <u>any</u> deformation of the probabilities $\{P_\theta(B)\}$ for $B \varepsilon B_X$ that leaves 0 and 1 probabilities invariant will not affect strong separation.

Now consider the various definitions of orthogonality of two classes of measures. It is clear that strong orthogonality implies weak orthogonality which implies pairwise orthogonality. Pairwise orthogonality does not imply weak orthogonality as is illustrated by example. Let $P_1 = \{\delta_{2\theta} : \theta \varepsilon [0,\frac{1}{2}]\}$ and let $P_2 \equiv \{\lambda\}$ where λ is Lebesgue measure on $[0,1]$. Clearly, $P_1 \perp (pr) P_2$, yet $\mu = \int \delta_{2\theta} d\theta$, an average of

elements of P_1, is Lebesgue measure on $[0,1]$, giving that P_1 and P_2 are not weakly orthogonal.

Under Assumption M, we have no examples of weakly separated families that are not strongly separated. Some of the results that are proven are stated below, and their proofs are contained in the following sections.

If P_1 and P_2 contain discrete measures only, then under Assumption M, pairwise orthogonality implies strong orthogonality. This is proved in Section 2 and Section 3.

If P_1 and P_2 consist of measures that are dominated by a single measure, then pairwise orthogonality implies strong orthogonality. Assumption M is not needed here, in fact, no measurability condition is needed at all. These facts are demonstrated in Section 4.

Let P_1 and P_2 consist of measures having a discrete part, and a continuous part which is dominated by a single sigma-finite measure. Then under Assumption M, weak orthogonality implies strong orthogonality. This is proved in Sections 5 and 6.

Let Q_γ, $\gamma \varepsilon [0,1]$, be the distribution of the random variable $Z = \sum_{k=1}^{\infty} 2^{-k} Y_k$ where $\{Y_k\}$ are independent Bernoulli random variables for which $P\{Y_k = 1\} = \gamma$. Call $S = \{Q_\gamma : \gamma \varepsilon [0,1]\}$ the coin tossing family of measures. The family S is an undominated family of continuous measures on $[0,1]$. In Section 7 it is shown that if $P_1, P_2 \subset S$ then under Assumption M, pairwise separation of P_1 with P_2 is equivalent to strong separation.

Section 2. The Most Basic Case

Let us first examine the case for which $P_1 \perp (pr) P_2$, and $P_\theta = \delta_{a(\theta)}$, that is, each measure P_θ is just the point mass measure at some point $a(\theta)$.

Lemma 2.1. *Under Assumption M, the function* $a: \Theta \to X$ *is a Borel measurable function of* θ.

Proof: Trivial since $\{\theta: a(\theta) \varepsilon B\} = \{\theta: P_\theta(B) = 1\}$ and if $B \varepsilon \mathcal{B}$ the second set is a Borel set of θ's by Assumption M. □

Proposition 2.2. *Under Assumption M if* $P_1 \perp (pr) P_2$ *and* $P_\theta = \delta_{a(\theta)}$, *then there exists a Borel subset* B *of* $[0,1]$, *such that* $P_\theta(B) = 1$ *if* $\theta \varepsilon \Theta_1$ *and* $P_\theta(B) = 0$ *if* $\theta \varepsilon \Theta_2$.

Proof: By Lemma 2.1, $a(\theta)$ is a measurable function, so that $a(\Theta_1)$ and $a(\Theta_2)$ are analytic sets. The sets $a(\Theta_1)$ and $a(\Theta_2)$ are disjoint since the classes P_1 and P_2 are separated by pairs. By the analytic set separation theorem (Kuratowski, Section 39), there is a Borel set B that contains $a(\Theta_1)$ and whose complement contains $a(\Theta_2)$. We will then have that $P_\theta(B) = 1$ or 0, according as $\theta \varepsilon \Theta_1$ or $\theta \varepsilon \Theta_2$. □

Section 3. The General Discrete Case

In this section, we will treat the case in which, for all θ, P_θ is a purely discrete measure. What is to be proved is that, under Assumption M, if P_1 and P_2 are separated by pairs then they are separated strongly.

In Section 2, each measure P_θ was a point mass at $a(\theta)$, which was shown to be a Borel function. A similar argument will work in this case. Rather than the one function $a(\theta)$, a countable number of functions $f_{n,k}(\theta)$ are needed. For each n, k, $f_{n,k}(\theta)$ will be one of the atoms of P_θ. (For some values of n, k and θ, $f_{n,k}(\theta)$ might be a special value to indicate that $f_{n,k}(\theta)$ is not an atom of P_θ. The reason for this will become clear below.)

The functions $f_{n,k}(\theta)$ are defined as follows. Consider $A_\theta^n = \{x \varepsilon [0,1]: P_\theta\{x\} \geq \frac{1}{n}\}$. If A_θ^n has at least k elements then define $f_{n,k}(\theta)$ to be the k^{th} largest element in A_θ^n. If A_θ^n does not have at least k elements, then define $f_{n,k}(\theta)$ to be minus one, the special element. The function $f_{n,k}(\theta)$ is then the k^{th} largest atom of P_θ having probability greater than or equal to $\frac{1}{n}$. (Largest is in the sense of the ordering on $X = [0,1]$, not the most probable.) As defined, $f_{n,k}$ is then a Borel function by

Lemma 3.1. Under Assumption M, for each $n = 1,2,\ldots$ and $k = 1,2,\ldots,n$, $f_{n,k}: \Theta \to [0,1] \cup \{-1\}$ is a Borel measurable function.

Proof: As is usual, it is enough to show for each fixed $x \varepsilon [0,1]$, that the set $C^{n,k} = \{\theta: f_{n,k}(\theta) \geq x\}$ is a Borel subset of Θ. Let n, k, x be fixed. Define the sets of θ:

$$B_{\ell,m}^n = \{\theta: P_\theta\{[\ell/2^m, (\ell+1)/2^m) \cap [x,1]\} \geq \frac{1}{n}$$

defined for $m = 1,2,\ldots$ and $\ell = 0,1,\ldots,2^m$,

$$B_m^{n,k} = \{\theta: \text{ there exist } \ell(1),\ell(2),\ldots,\ell(k) \text{ such that } 0 \le \ell(1) < \ell(2) < \ldots < \ell(k) \le 2^m \text{ and that } \theta \in \cap_{i=1}^{k} B^n_{\ell(i),m}\}$$

defined for $m = 1,2,\ldots,$

$$B^{n,k} = \{\theta: \text{ there exists } M \ge 1 \text{ such that } \theta \in \cap_{m=M}^{\infty} B_m^{n,k}\} .$$

The sets $B^n_{\ell,m}$ are Borel sets by Assumption M. The sets $B_m^{n,k}$, $B^{n,k}$ will also be Borel sets since they are obtained through countable intersections and unions of sets $B^n_{\ell,m}$. We clain that $C^{n,k} = B^{n,k}$. This is true because $P_\theta\{y\} \ge \frac{1}{n}$ if and only if all small intervals I_y containing y have probability of at least $\frac{1}{n}$. The set $B_m^{n,k}$ is the set of θ for which at least k intervals of the form $[\ell/2^m,(\ell+1)/2^m)$ have P_θ probability at least $\frac{1}{n}$. Clearly, in order that $f_{n,k}(\theta) \ge x$ it is necessary and sufficient that $B_m^{n,k}$ occur all but finitely often. Thus $C^{n,k} = B^{n,k}$ as claimed, proving that for all n, k, $f_{n,k}(\cdot)$ is a Borel measurable function. $\square$

Remark: The proof of the lemma nowhere uses the fact that the P 's are purely discrete; only Assumption M is used. We will make use of this fact in Section 6.

We are now in a position to prove the following.

Theorem 3.1. If for all θ, P_θ is a purely discrete measure, Assumption M holds, and $P_1 \perp (pr) P_2$ then $P_1 \perp (s) P_2$

Proof: Since $f_{n,k}$ is a measurable function, the sets $f_{n,k}(\Theta_1)$, $f_{n,k}(\Theta_2)$ are analytic sets by the definition of analytic set. Let $A_1 = \cup_{n,k} f_{n,k}(\Theta_1) \setminus \{-1\}$ and $A_2 = \cup_{n,k} f_{n,k}(\Theta_2) \setminus \{-1\}$. The sets A_1 and A_2 are then analytic since the countable union of analytic sets is an analytic set. (See Kuratowski.) A_1, A_2 must also be disjoint, since $P_1 \perp (pr) P_2$, forcing P_1, P_2 to have no atoms in common. We now can apply the analytic set separation theorem as in Section 2 to get a Borel set B that contains A_1, and whose complement contains A_2. The set A_1 contains all atoms of measures in P_1, since if $P_\theta\{x\} > 0$ then $P_\theta\{x\} \geq \frac{1}{n'}$ for some n' giving $f_{n',k}(\theta) = x$ for some $k' \leq n'$. We then have that $P_\theta(B) = 1$ for $\theta \in \Theta_1$ and $P_\theta(B) = 0$ for $\theta \in \Theta_2$. □

Remark: There is a slightly cleaner proof of this theorem not using Lemma 3.1 -- however the proof given above lends itself directly to Section 5 below, whereas the other proof does not. The alternate proof is this: The set $A^* = \{(x,\theta): P_\theta\{x\} > 0, \ \theta \in \Theta_1\}$ is a Borel set in the plane. The set, $\pi_X(A^*)$, the projection of A^* onto $X = [0,1]$, is then an analytic set. As above, apply the analytic set separation theorem to yield the desired result.

Section 4. The Case of Dominated Measures

In this section, we will deal with the situation in which the collection $\{P_\theta\}$ is dominated by some sigma-finite measure λ, that is, $P_\theta << \lambda$ for all θ.

Let M_1, M_2 be two collections of sigma-finite measures on a measurable space X. We say that M_1 and M_2 are equivalent if: $m_1(E) = 0$ for all m_1 in M_1 implies $m_2(E) = 0$ for all $m_2 \in M_2$, and conversely. The following lemma of Halmos and Savage (1949) allows us to treat the dominated case.

Lemma 4.1. Every dominated set of measures has an equivalent countable subset.

Proof: Refer to Lemma 7 of Halmos and Savage. □

The following easy result is also useful.

Lemma 4.2. Let P, R be two countable sets of sigma-finite measures on some measurable space, then if P and R are separated by pairs, they are also separated strongly.

Proof: Say that $P = \{P_1, P_2, \ldots\}$ and $R = \{R_1, R_2, \ldots\}$. For each i, j, $P_i \perp R_j$ by assumption, so there exists a Borel set B_{ij} such that $P_i(B_{ij}^c) = 0$ and $R_j(B_{ij}) = 0$. If we define the Borel set B to be $\cup_i \cap_j B_{ij}$, then for all i and j, $P_i(B^c) = 0$ and $R_j(B) = 0$. □

We can now tie these two lemmas together to get

Theorem 4.1. Let $P_1 \perp (pr) P_2$ and let there be a sigma-finite measure λ such that $P_\theta \ll \lambda$ for all $\theta \in [0,1]$, then $P_1 \perp (s) P_2$.

Proof: By Lemma 4.1, P_1 and P_2 each has a countable equivalent subset. Now, use Lemma 4.2 on the countable equivalent subsets of P_1 and P_2 to obtain the desired result. □

Remark: Assumption M was not required at all; only the pairwise separation of the two classes was needed. Assuming a class to be dominated is a very strong assumption.

Up to this point we have not needed the assumption that the classes P_1 and P_2 be weakly separately, just pairwise separation was used. In the following section we will require the weak separation assumption.

Section 5. An Arbitrary Measure Versus a Class of Discrete Measures.

What has been treated so far are the cases in which either all measures are discrete, or all measures are absolutely continuous with respect to a single measure (such as Lebesgue measure). In this section we shall examine the situation of one continuous measure in contrast to a collection of purely discrete measures.

Precisely what is assumed is that $P_1 \equiv \{Q\}$ for some probability Q , and P_2 consists of purely discrete measures. As we know from Section 3, $A = \{\text{atoms of measures in } P_2\}$ is an analytic set. While

the set A need not be a Borel set, it is measurable with respect to Q, more precisely with respect to the completion of the measure Q.

The completion of a measure μ will be denoted as $\bar{\mu}$. The quantity $\bar{Q}(A)$ is then well defined. We are led to the following proposition.

Proposition 5.1. If

(1) Assumption M holds,

(2) $\mathcal{P}_1 \equiv \{Q\}$, $\mathcal{P}_2$ are all purely discrete,

(3) $\bar{Q}\{\mathcal{P}_2 \text{ atoms}\} = 0$

then the families $\mathcal{P}_1$ and $\mathcal{P}_2$ are strongly separated.

Proof: Let $A = \{\text{atoms of measures in } \mathcal{P}_2\}$. The set A is analytic so there exist Borel sets B_1, B_2 such that $B_1 \subset A \subset B_2$ and $Q(B_2) = Q(B_1)$. But $Q(B_2) = 0$ by assumption, giving that $P(B_2) = 0$ for $\theta \varepsilon \Theta_1$, and $P(B_2) = 1$ for $\theta \varepsilon \Theta_2$. (That is, $\mathcal{P}_1$ and $\mathcal{P}_2$ are strongly separated families of probability measures).

□

What remains is the case in which $\bar{Q}(A) > 0$. If $\bar{Q}(A) > 0$ then it is not possible that the families $\mathcal{P}_1$ and $\mathcal{P}_2$ can be weakly separated as will be shown below. This is accomplished by finding an appropriate probability measure ν on Θ_2. The measure $\mu = \int P_\theta \nu(d\theta)$ will be an average of measures in $\mathcal{P}_2$ yet not orthogonal to the measure Q, giving a contradiction of weak separation.

In order to find the desired measure ν, the first step is to find a "nice" subset of the set of atoms. This is facilitated by the following

Lemma 5.1. Under Assumption M, if $\bar{Q}\{\text{atoms}\} > 0$ then there exist n', k' such that $\bar{Q}\{f_{n',k'}(\Theta_2)\} > 0$.

Proof: Follows directly from the proof of Theorem 3.1 and the measurability of analytic sets. $\square$

From this point on, assume that some n', k' have been chosen so $\bar{Q}\{f_{n',k'}(\Theta_2)\} > 0$. Denote the function $f_{n',k'}$ by f.

A *heuristic argument of what is to be attempted* follows. What we would like to do is to construct ν from Q and f in a way that would give $\nu(f^{-1}(\cdot)) = Q(\cdot)$. This would identify null sets of Q with those of ν. If $\mu(\cdot)$ were defined as $\int P_\theta(\cdot)\nu(d\theta)$ then null sets of Q would have μ-probability less than one. This would be true since for N a Q-null set, ν assigns probability less than one to N. This argument will be rigorized below.

However, this approach cannot be used as it stands since ν is only defined on a sub sigma-field of the Borel sigma-field, namely that generated by the measurable function f. What follows basically is a means of extending the definition of ν to the entire Borel sigma-field.

The paper "Borel Structure in Groups and Their Duals" by George Mackey (1957) contains the needed results. He defines a Borel space to be standard whenever it is Borel isomorphic to the Borel space associated with a Borel subset of a complete separable metric space. A sigma-finite measure λ on a Borel space S is standard if there exists a Borel subset E of S so that E regarded as a subspace of S is standard and that $\lambda(S \setminus E) = 0$. In this paper, all Borel spaces and all measures used are standard, by definition.

The theorem below is a direct statement of Mackey's Theorem 6.3 in the paper mentioned above.

Theorem 5.1. *Let* S_1 *and* S_2 *be Borel spaces and let* S_2 *be stanard. Let* μ *be a standard measure in* S_1. *Let* A *be a Borel subset of* $S_1 \times S_2$ *such that for each* $x \in S_1$ *there exists* $y \in S_2$ *so that* $(x,y) \in A$. *Then there exists a Borel subset* N *of* S_1 *and a Borel function* ϕ *from* $S_1 \setminus N$ *to* S_2 *such that* $(x,\phi(x)) \in A$ *for all* $x \in S_1 \setminus N$ *and* $\mu(N) = 0$.

This theorem and some of its ramifications is treated in detail in the monograph of T. Parthasarathy, (1971), *Selection Theorems and Their Applications*.

We can apply the theorem in the following way. The graph $G = (f(\theta),\theta)$ of the function f is a Borel set in the product space $X \times \Theta_2$. Then $f(\Theta_2)$ is $\pi_X(G)$ for π_X the projection map onto X. By the assumptions above, there exists a Borel set B^* contained in $f(\Theta_2)$ such that

$Q(B^*) > 0$. The set $G \cap \pi_X^{-1}(B^*)$ will be a Borel set in the product space which will serve as the set A of the theorem above.

The application of the theorem is stated as

Lemma 5.2. For f and Q defined as in this section, there exists a Borel subset B' of $f(\Theta_2)$ with $Q(B') > 0$, and a Borel measurable function ϕ mapping $B' \to \Theta_2$ such that for all $x \in B'$, $(x,\phi(x)) \in \{(f(\theta),\theta): \theta \in \Theta_2\}$.

Proof: Let $A = G \cap \pi_X^{-1}(B^*)$ which is a Borel subset of $X \times \Theta$. The set B^* is a subset of $f(\Theta_2)$ so that π_X maps A to all of B^* . The set B^* can serve as the Borel space S_1 of Theorem 5.1. The set Θ_2 serves as S_2 . The theorem asserts the existence of a Q-null set N and a Borel function ϕ mapping $B^* \setminus N$ to Θ_2 such that $(x,\phi(x)) \in G$ for all $x \in B^* \setminus N$. Then the set $B' = B^* \setminus N$ is the desired set and the function ϕ is the desired function. □

Remark: The construction of ϕ gives us that ϕ is a one-one function. As is demonstrated below ϕ is a one sided inverse of the function f .

Define the measures ν on Θ_2 and μ on X as

(5.1) $$\nu(\cdot) = \frac{1}{Q(B')} Q(B' \cap \phi^{-1}(\cdot))$$

and

(5.2) $$\mu(\cdot) = \int P_\theta(\cdot)\nu(d\theta) .$$

That μ cannot be orthogonal to Q is demonstrated below in the proof of

Proposition 5.2. If

(1) Assumption M holds,

(2) $P_1 \equiv \{Q\}$, P_2 are all purely discrete,

(3) $\bar{Q}\{P_2 \text{ atoms}\} > 0$

then P_1 and P_2 cannot be weakly separated.

Proof: Let μ, ν be defined as in (5.1) and (5.2). We want to show that $C \in \mathcal{B}_X$ and $Q(C) = 1$ implies $\mu(C) > 0$, which would demonstrate that μ is not orthogonal to Q . First recall that f is $f_{n',k'}$ for some n', k' . The function f has as its image, the point $\{-1\}$ and atoms of P_θ whose P_θ probability is at least $\frac{1}{n'}$. We then get that for any $C \in \mathcal{B}_X$ (denoting $\frac{1}{n'}$ as ε)

$$(5.3) \qquad P_\theta(C) \geq \varepsilon I_C(f(\theta)) .$$

(The element $-1 \notin X$, and $C \in \mathcal{B}_X$ so -1 is not an element of C.) So we have

$$(5.4) \qquad \mu(C) \geq \varepsilon\nu\{\theta\colon f(\theta) \in C\} = \frac{\varepsilon}{Q(B')} Q(B' \cap \phi^{-1}\{\theta\colon f(\theta) \in C\}) .$$

Now note, if $y \in B' \cap C$ then, since $B' \subset f(\Theta_2)$ there is at least one $\theta' \in \Theta_2$ for which $f(\theta') = y$. By the construction of ϕ , $\phi(y) \in \{\theta\colon f(\theta) = y\}$, so that if $y \in B' \cap C$ then $y \in B' \cap \phi^{-1}\{\theta\colon f(\theta) \in C\}$. Referring back to (5.4), for any $C \in \mathcal{B}_X$ the given inequality holds:

$$\mu(C) \geq \frac{\varepsilon}{Q(B')} Q(B' \cap C) .$$

Suppose that $Q(C) = 1$, then clearly

$$\mu(C) \geq \varepsilon .$$

But $\varepsilon > 0$ so that μ cannot be orthogonal to Q . Since μ is an integral average of measures in P_2 , we have shown that the families P_1 and P_2 are not weakly orthogonal. □

Propositions 5.1 and 5.2 lead directly to the desired fact:

Theorem 5.2. If

(1) Assumption M holds,

(2) $P_1 \equiv \{Q\}$, P_2 consists only of discrete measures,

then if P_1 and P_2 are weakly orthogonal, they are also strongly orthogonal.

Proof: Proposition 5.2 gives that if P_1 and P_2 are weakly orthogonal then $\bar{Q}\{P_2 \text{ atoms}\} = 0$. Proposition 5.1 asserts that P_1 and P_2 must then be strongly orthogonal. □

This section is now complete. In the next section, all preceding results are combined into one theorem.

Section 6. Mixed Measures with a Dominated Continuous Part

In Section 3 all measures considered were purely discrete. In Section 4 all measures were assumed to be dominated by a single sigma finite measure. The last section dealt with the case of a single

continuous measure versus a class of purely discrete measures. Now we wish to combine all the above cases into one.

Every probability measure can be uniquely decomposed into the sum of two sub-probability measures, one of which is discrete and the other purely continuous. We then have that for each θ, $P_\theta = P^a_\theta + P^c_\theta$. The measure P^a_θ is purely discrete (atomic) and the measure P^c_θ is purely continuous.

What is to be assumed in the rest of this section is there exists a sigma-finite measure λ such that for all θ, $P^c_\theta << \lambda$.

The following notation will be useful. Define

$$\mathcal{P}^a_1 = \{P^a_\theta : \theta \varepsilon \Theta_1\}\ , \qquad \mathcal{P}^a_2 = \{P^a_\theta : \theta \varepsilon \Theta_2\}$$

$$\mathcal{P}^c_1 = \{P^c_\theta : \theta \varepsilon \Theta_1\}\ , \qquad \mathcal{P}^c_2 = \{P^c_\theta : \theta \varepsilon \Theta_2\}\ .$$

The orthogonality of $\mathcal{P}_1$ and $\mathcal{P}_2$ is equivalent to the orthogonality of $\mathcal{P}^a_1$ with $\mathcal{P}^a_2$, $\mathcal{P}^c_1$ with $\mathcal{P}^c_2$, $\mathcal{P}^a_1$ with $\mathcal{P}^c_2$, and $\mathcal{P}^a_2$ with $\mathcal{P}^c_1$. Each of these cases will be considered one at a time starting with $\mathcal{P}^a_1$ versus $\mathcal{P}^a_2$.

The lemma below is derived in Section 3.

Lemma 6.1. If A_1 = {atoms of measures in $\mathcal{P}_1$} and A_2 = {atoms of measures in $\mathcal{P}_2$} then A_1 and A_2 are disjoint analytic sets.

Proof: Refer to the proof of Theorem 3.1. □

The next case to be dealt with is that of $\mathcal{P}^c_1$ versus $\mathcal{P}^c_2$. By Lemma 4.1, the classes of $\mathcal{P}^c_1$ and $\mathcal{P}^c_2$ each has a countable equivalent

subset. (Whether the measures involved are probability measures or subprobability measures does not matter.) Enumerate the countable equivalent subsets (ces) of $\mathcal{P}_1^c$ and $\mathcal{P}_2^c$ as $\mathrm{ces}(\mathcal{P}_1^c) = \{P_{\theta_1}^c, P_{\theta_2}^c, \ldots\}$ and $\mathrm{ces}(\mathcal{P}_2^c) = \{P_{\tau_1}^c, P_{\tau_2}^c, \ldots\}$.

We are now ready to prove the major result of this section.

Theorem 6.1. If

(1) Assumption M holds,

(2) there exists a sigma-finite measure λ, such that for all θ, $P_\theta^c \ll \lambda$,

then if $\mathcal{P}_1$ and $\mathcal{P}_2$ are weakly orthogonal, they are also strongly orthogonal.

Proof: Let $P_{\theta_i}^c \in \mathrm{ces}(\mathcal{P}_1^c)$ and let $P_{\tau_j}^c \in \mathrm{ces}(\mathcal{P}_2^c)$. Since $\mathcal{P}_1 \perp (w) \mathcal{P}_2$, it follows directly that $P_{\theta_i}^c \perp P_{\tau_j}^c$. There is a Borel set B_{ij} for which

$$P_{\theta_i}^c(B_{ij}^c) = 0 \quad \text{and} \quad P_{\tau_j}^c(B_{ij}) = 0 .$$

In order to get the strong separation of $\mathcal{P}_1$ and $\mathcal{P}_2$ it is desirable that $B_{ij} \supset A_1$ and $B_{ij} \subset A_2^c$, for A_1 the set of atoms of measures in $\mathcal{P}_1$ and A_2 the set of atoms of measures in $\mathcal{P}_2$. This may not be so for arbitrary B_{ij}.

The results of Section 5 lead to the fact that for all $\theta \in \Theta_1$, $\bar{P}_\theta(A_2) = 0$ and for all $\theta \in \Theta_2$, $\bar{P}_\theta(A_1) = 0$. Because $\theta_i \in \Theta_1$, $\bar{P}_{\theta_i}(A_2) = 0$, so there is a Borel set $C_i \supset A_2$ for which $P_{\theta_i}(C_i) = 0$. Because $\tau_j \in \Theta_2$, $\bar{P}_{\tau_j}(A_1) = 0$, so there is a Borel set $D_j \supset A_1$ for which $P_{\tau_j}(D_j) = 0$. The set $E_{ij} = B_{ij} \cup D_j \cap C_i^c$ has the properties

that $E_{ij} \supset A_1$, $E_{ij} \subset A_2^c$, and $P_{\theta_i}(E_{ij}) = 1$ and $P_{\tau_j}(E_{ij}) = 0$.

As in the proof of Lemma 4.2, let $E = \cup_i \cap_j E_{ij}$. The set E contains A_1, is contained in A_2^c, and for all i, j, $P_{\theta_i}(E) = 1$ and $P_{\theta_j}(E) = 0$. Since E contains A_1, $P_\theta^a(E^c) = 0$ for all $\theta \varepsilon \Theta_1$. The set E is contained in A_2^c so $P_\theta^a(E) = 0$ for all $\theta \varepsilon \Theta_2$. We also have $P_\theta(E^c) = 0$ if θ is a θ_i but the definition of countable equivalent subset forces $P_\theta(E^c) = 0$ for <u>all</u> $\theta \varepsilon \Theta_1$. Similarly $P_\theta(E) = 0$ for <u>all</u> $\theta \varepsilon \Theta_2$.

What has been shown is that $P_\theta(E) = 1$ for all $\theta \varepsilon \Theta_1$ and $P_\theta(E) = 0$ for all $\theta \varepsilon \Theta_2$. The families $\mathcal{P}_1$ and $\mathcal{P}_2$ are separated strongly if they are separated weakly. □

The results of this section can be extended somewhat. Following is an extension that gives an answer in some cases not covered so far, particularly that of certain undominated families of continuous measures.

Section 7. Discrete-Like Undominated Families of Measures

Consider the family of coin tossing measures mentioned in Section 1. It was stated there that if $\mathcal{P}_1$ and $\mathcal{P}_2$ consist of such measures, then weak and strong orthogonality are equivalent. The coin tossing family is an example of a class of families of measures, that we will call Discrete-Like Undominated Families (DLUF for short).

First is a definition of DLUF's, followed by the proof of the fact that weak and strong separation are equivalent for DLUF's. At the end of this section, we will prove the easy fact that the coin-tossing family is a DLUF.

The family of measures $\mathcal{S} = \{Q_\gamma\}$ for $\gamma \varepsilon \Gamma = [0,1]$ is said to be a Discrete-Like Undominated Family if there exists a Borel function

$f: X \to \Gamma$ that partitions X into disjoint Borel sets in a measurable way. By this is meant that f is Borel (with respect to the usual Borel sigma-fields on X and Γ), and $Q_\gamma(f^{-1}(\gamma)) = 1$ and $Q_\gamma(f^{-1}(\gamma')) = 0$ if $\gamma \neq \gamma'$.

Suppose now that $S = \{Q_\gamma\}$ is some DLUF and $P \subset S$. Let $g: \Theta \to \Gamma$ be the function that identifies each θ with its corresponding γ-atom, that is, for each $\theta \in \Theta$, $P_\theta = Q_{g(\theta)}$. The function $g(\cdot)$ is analogous to the function $a(\cdot)$ of Section 2.

Lemma 7.1. If S is a DLUF, $P \subset S$ and Assumption M holds, then the function $g: \Theta \to \Gamma$ as defined above is a Borel function.

Proof: This lemma is a generalization of Lemma 2.1, its proof is almost identical. Let B be a Borel subset of Γ, then $\{\theta: g(\theta) \in B\} \subset \{\theta: P_\theta(f^{-1}(B)) = 1\}$, since $g(\theta) \in B$ implies $Q_{g(\theta)}(f^{-1}(B)) = 1$ by the identification of f; but $P_\theta = Q_{g(\theta)}$ so that $P_\theta(f^{-1}(B)) = 1$. The relation $\{\theta: g(\theta) \in B\} \supset \{\theta: P_\theta(f^{-1}(B)) = 1\}$ holds, because if $P_\theta(f^{-1}(B)) = 1$ then $Q_{g(\theta)}(f^{-1}(B)) = 1$ forcing $g(\theta) \in B$ due to the "partitioning" nature of the function f. Since $\{\theta: P_\theta(f^{-1}(B)) = 1\}$ is a Borel set by Assumption M, g is a Borel function of θ. □

Using the analytic set separation theorem as in Section 2 we get

Proposition 7.1. If $P \subset S$, for S some DLUF then under Assumption M, the pairwise separation of the families P_1 and P_2 implies the strong separation of P_1 and P_2.

Proof: Lemma 7.1 together with pairwise separation gives us that the sets $A_1 = g(\Theta_1)$ and $A_2 = g(\Theta_2)$ are disjoint analytic subsets of Γ. By the analytic separation theorem, there exists B, a

Borel subset of Γ which contains A_1 and whose complement contains A_2. The set $C = f^{-1}(B)$ is clearly a Borel subset of X such that $P_\theta(C) = 1$ for $\theta \in \Theta_1$ and $P_\theta(C) = 0$ for $\theta \in \Theta_2$. □

The family S is similar to a collection of one-point measures except that Q_γ is not point mass at γ but rather like "point-mass" at the "point" $f^{-1}(\gamma)$. The generalization to measures that are convex combinations of measures in S leads to a notation of discrete measures whose "atoms" are of the form $f^{-1}(\gamma)$.

Define $S^* = \{\sum_{i=1}^{\infty} \alpha_i Q_{\gamma_i} : \alpha_i \geq 0 , \sum \alpha_i = 1 , \gamma_i \in \Gamma$ and $Q_\gamma \in S\}$, set of all countably infinite convex combinations of elements of S. We want to show that if $P \subset S^*$ then weak orthogonality of P_1 with P_2 implies strong orthogonality. This is accomplished by proceeding just as in Section 3.

Let $P \subset S^*$. We will define $g_{n,k} : \Theta \to \Gamma$ similarly to $f_{n,k}$. Let $A_\theta^n = \{\gamma \in \Gamma = [0,1]: P_\theta\{f^{-1}(\gamma)\} \geq \frac{1}{n}\}$. If A_θ^n has at least k elements then define $g_{n,k}(\theta)$ to be the k^{th} largest element of A_θ^n. If A_θ^n does not have at least k elements then define $g_{n,k}(\theta)$ to be minus one, a special element. By examining the proof of Lemma 3.1, it is easily seen for $n = 1,2,\ldots$ and $k = 1,\ldots,n$ that the function $g_{n,k}(\cdot)$ is measurable if Assumption M holds.

<u>Proposition 7.2</u> <u>If</u> S <u>is a DLUF,</u> $P \subset S^*$, <u>and Assumption</u> M <u>holds, then</u> $P_1 \perp (pr) P_2$ <u>implies</u> $P_1 \perp (s) P_2$.

<u>Proof</u>: As stated above, the function $g_{n,k}$ is a Borel function for each n, k. The sets $A_1 = \cup_{n,k}\, g_{n,k}(\Theta_1) \setminus \{-1\}$ and

$A_2 = \cup_{n,k}\, g_{n,k}(\Theta_2)\setminus\{-1\}$ are disjoint analytic subsets of Γ. The sets A_1 and A_2 consist of the γ-atoms for the measures in P_1, and P_2 respectively. There is a Borel set B such that $B \supset A_1$ and $B \cap A_2^c = \emptyset$. If $C = f^{-1}(B)$ then as before C will separate the families P_1 and P_2. □

If P consists of measures of the form $P_\theta = \int Q_\gamma \nu_\theta(d\gamma)$ and the measures ν_θ are dominated by some λ, then pairwise separation gives strong separation of P_1 and P_2.

The following theorem is an obvious generalization of Theorem 6.1.

<u>Theorem 7.1.</u> <u>Let</u> S <u>be some DLUF and</u> $S^\star$ <u>the set of all countably infinite convex combinations of members of</u> S. <u>Let</u>

(1) <u>Assumption M hold</u>,

(2) $P_\theta = \alpha(\theta)\int Q_\gamma \nu_\theta(d\gamma) + (1-\alpha(\theta))Q_\theta^\star$ <u>for</u> $0 \le \alpha(\theta) \le 1$, $\nu_\theta << \lambda$ <u>and</u> $Q_\theta^\star \in S^\star$,

<u>then if</u> P_1 <u>and</u> P_2 <u>are weakly orthogonal, they are also strongly orthogonal</u>.

<u>Proof:</u> The only case that has not been resolved in this section that was considered before is of $P_1 \equiv \{\int Q_\gamma \nu(d\gamma)\}$ and $P_2 \subset S^\star$. But this is like the case of $P_1 \equiv \{\nu\}$ and P_2 a set of discrete measures. As before the set of γ-atoms of P_2 measures is an analytic set for which $\bar{\nu}\{\text{atoms}\} = 0$. There is a Borel set B in Γ that contains the γ atoms and $\nu(B) = 0$. The set $f^{-1}(B)$ will split the single continuous measure from the discrete ones. Now go on as before to get the result. □

Two examples of DLUF's are the atomic masses and the coin tossing measures.

If $\mathcal{S} = \{\delta_\gamma : \gamma \varepsilon [0,1]\}$ then $\mathcal{S}$ is a DLUF since $f: X \to \Gamma$ for $f(x) = x$ satisfies the requirements stated above. It follows then that all the results of Sections 2 through 6 are special cases of the Theorem 7.1.

If $\mathcal{S} = \{Q_\gamma : \gamma \varepsilon [0,1]\}$ for Q_γ the coin tossing measure with probability γ then $\mathcal{S}$ is a DLUF. The measure Q_γ is the distribution of $Y_\gamma = \sum_{k=1}^{\infty} \frac{X_k}{2^k}$ where the X_k are i.i.d. Bernoulli random variables with parameter γ. Express $x \varepsilon X = [0,1]$ in terms of its binary expansion $x = \sum_{k=1}^{\infty} b_k/2^k$ and if x has two expansions, use the one with non-terminating ones. The function $f(x) = \limsup_{n\to\infty}(\frac{b_1+\ldots+b_n}{n})$ is a Borel function, being the countable lim sup of Borel functions. By the strong law of large numbers, $Q_\gamma(f^{-1}(\gamma)) = 1$ and $Q_\gamma(f^{-1}(\gamma')) = 0$ for $\gamma \neq \gamma'$. The family $\mathcal{S}$ is then a Discrete-Like Undominated Family.

Section 8. Conclusion

In this paper, it has been shown that for families usually considered in the study of hypothesis testing, weak orthogonality implies strong orthogonality. For discrete families, dominated continuous families, and for mixtures of discrete measures with dominated continuous measures, the two notions of orthogonality are equivalent.

At this time, the author does not know of any counterexample. Nor does he know of any general proof of the equivalence of weak and strong separation, if there is such a proof.

Acknowledgements

This paper is taken from the author's Ph.D. dissertation which was prepared at the University of California, Berkeley. The author wishes to express his gratitude to David Blackwell for suggesting this problem and for his help and encouragement throughout the period of research. In addition, the author wishes to thank Lucien LeCam for a couple of helpful conversations on this problem.

References

1. DUBINS, Lester and FREEDMAN, David (1964). "Measurable Sets of Measures", Pacific Journal of Math. **14**, 1211-1222.
2. HALMOS, Paul R. and SAVAGE, L. J. (1949). "Application of the Radon-Nikodym Theorem to the Theory of Sufficient Statistics," Annals of Mathematical Statistics **20**, 225-241.
3. KURATOWSKI, K (1966). Topology, Volume 1. Jointly published: Academic Press and PWN-Polish Scientific Publishers.
4. MACKEY, George W. (1957). "Borel Structure in Groups and Their Duals," Trans. Amer. Math. Soc. **85**, 134-165.
5. PARTHASARATHY, K. (1967). Probability Measures on Metric Spaces. New York: Academic Press.
6. PARTHASARATHY, T. (1971). Selection Theorems and Their Applications. Springer-Verlag Lectures Notes in Mathematics, Vol. 263.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 6	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) On the Strict Determination of Hypothesis Testing Games		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Richard Clark Persons		8. CONTRACT OR GRANT NUMBER(s) N00014-75-C-0555
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Mathematics M.I.T., Cambridge, Massachusetts 02139		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS (NR-042-331)
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Statistics & Probability Program Code 436 Arlington, Virginia 22217		12. REPORT DATE July 22, 1977
		13. NUMBER OF PAGES 27
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for Public Release; Distribution Unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Hypothesis testing, orthogonal classes of measures, strictly determined games, Bayesian

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

(see reverse side)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Measurably parameterized subsets of the set of all probability measures on a space are examined. The setup involves an arbitrary composite null hypothesis and the complementary alternative hypothesis. Suppose that for every probability measure on the parameter space, the averaged null hypothesis measure is orthogonal to the averaged alternative hypothesis measure. Whether, there then exists a Borel set in the range space which separates the null and alternative hypotheses is explored. It is shown that, in cases usually considered in hypothesis testing, there does exist such a Borel set. The relevance to this question to Bayesian vs. classical statistics is also discussed.

Measurably parameterized subsets of the set of all probability measures on a space are examined. The setup involves an arbitrary composite null hypothesis and the complementary alternative hypothesis. Suppose that for every probability measure on the parameter space, the averaged null hypothesis measure is orthogonal to the averaged alternative hypothesis measure. Whether, there then exists a Borel set in the range space which separates the null and alternative hypotheses is explored. It is shown that, in cases usually considered in hypothesis testing, there does exist such a Borel set. The relevance to this question to Bayesian vs. classical statistics is also discussed.